

8. Schmidt CW. Bordering on environmental disaster. *Environ Health Perspect* 200;108:308-15.
9. Shalat SL, Donnelly KC, Freeman NC, Calvin JA, Ramesh S, Jimenez M, et al. Nondietary ingestion of pesticides by children in an agricultural community on the U.S./Mexico border: preliminary results. *J Expo Anal Environ Epidemiol* 2003;13:42-50.
10. Carrillo-Zuniga G, Coutino C, Shalat SL, Freeman NCG, Black K, Jimenez W, et al. Potential sources of childhood exposure to pesticides in an agricultural community. *J Child Health* 2004;2:29-39.
11. Sumaya C, Carrillo-Zuniga G, Kelley M, May M, Zhu L, Donnelly KC. Linking research to health promotion in Texas colonias. *Am J Health Studies* 2006;21:245-52.

## On Academics

### OPENEPI: A WEB-BASED EPIDEMIOLOGIC AND STATISTICAL CALCULATOR FOR PUBLIC HEALTH

KEVIN M. SULLIVAN, PhD, MPH, MHA  
ANDREW DEAN, MD, MPH  
MINN MINN SOE, MD, MCTM, MPH

During graduate training in public health, students are exposed to a number of epidemiologic and statistical formulae. Frequently, many of these may be initially calculated by hand with a calculator and then later using computer software. A number of commercial programs are available, many of which are taught in schools of public health (SPHs), such as SAS®, SPSS®, and Stata®. While they are probably the most commonly used commercial programs in epidemiology,<sup>1</sup> and are incredibly powerful, they do have certain characteristics that limit their use in low- to moderate-resource situations: (1) they tend to be expensive beyond the reach of some SPHs worldwide, (2) they tend to be complicated to learn and use, (3) they have a number of hardware/computer requirements, and (4) user support may not be readily available.

Free epidemiologic programs, such as Epi Info ([www.cdc.gov/epiinfo](http://www.cdc.gov/epiinfo)) and EpiData ([www.epidata.dk](http://www.epidata.dk)), are also available. For some of the commercial and free programs mentioned, many epidemiologic parameters and statistical calculations taught in SPH courses are not provided in the programs. Because of the aforementioned limitations, we decided to develop a program to assist in teaching epidemiologic and statistical procedures.

#### INTRODUCTION TO OPENEPI

OpenEpi ([www.OpenEpi.com](http://www.OpenEpi.com)) is a free, Web-based, open-source, operating system-independent series of programs designed for use in public health and medicine—for training or practice—that provide a

number of epidemiologic and statistical tools for summary data.<sup>2-7</sup> OpenEpi was developed in JavaScript and hypertext markup language (HTML) and can be run in browsers supporting these languages. Therefore, it can run under various Web browsers, such as Microsoft® Explorer, Firefox®, Safari, and Opera, and on a number of operating systems, such as Windows, Macintosh, and Linux. It can even run on an iPhone. The program can be run from the OpenEpi website or downloaded and run without a Web connection. The source code and documentation are freely downloadable and available for use by other investigators. The program is available in English, French, Italian, and Spanish.

The OpenEpi developers have had extensive experience in developing and testing Epi Info, a program developed by the Centers for Disease Control and Prevention (CDC) and used worldwide for data entry and analysis. OpenEpi was initially developed to perform analyses found in the disk operating system (DOS) version of Epi Info modules StatCalc and EpiTable, to improve on the types of analyses provided by these modules, and to provide a number of tools and calculations not currently available in Epi Info. It is the first step toward an entirely Web-based suite of epidemiologic software tools. We see OpenEpi as a useful companion to Epi Info and to other programs such as SAS, SPSS, Stata, and EpiData. OpenEpi, Epi Info, and EpiData were developed with the goal of providing simple tools for low- and moderate-resource areas of the world.

The initial development of OpenEpi was supported by a grant from the Bill and Melinda Gates Foundation to Emory University. OpenEpi has had more than 1.4 million hits from 160 countries since its inception in 2003 and almost 500,000 hits in the first six months of 2008. A recent Google search identified almost 5,000 sites mentioning OpenEpi.

#### ANALYTIC CAPABILITIES OF OPENEPI

The types of calculations currently performed by OpenEpi are shown in Figure 1 and include:

- Various confidence intervals for proportions, rates, standard mortality ratio, mean, median, and percentiles

**Figure 1. OpenEpi menu with short descriptions of the modules**

Menu item	Description
Home	The main OpenEpi page
Info and Help	
Language/Options/Settings	Optional settings such as language and table orientation
Calculator	A calculator, similar to a handheld scientific calculator
Counts	<i>The following modules are for analysis of count data:</i>
Std.Mort.Ratio	SMR: standardized mortality ratio
Proportion	Various confidence intervals for a proportion
Two by Two Table	Two-by-two tables, crude and stratified two-by-two tables with <i>p</i> -values and estimates
Dose-Response	Tables for more than two exposure levels
R by C Table	General R by C table, Chi-square <i>p</i> -value
Matched Case Control	Pair-matched case-control analysis; <i>p</i> -values and odds ratios
Screening	Screening/diagnostic results such as sensitivity and specificity
Person Time	<i>The following modules are for analysis of person-time data:</i>
1 Rate	Various confidence intervals for a single rate
Compare 2 Rates	Two-by-two tables and stratified tables with person-time data
Continuous Variables	<i>The following modules are for analysis of continuous data:</i>
Mean CI	Provides a confidence interval for a mean
Median/%ile CI	Provides observations for a percentile and its confidence limits
t test	Calculates independent t-test assuming equal and unequal variance
ANOVA	Calculates analysis of variance table
Sample Size	<i>Sample size calculations for various study designs:</i>
Proportion	Absolute precision approach to sample size for a proportion
Unmatched CC	Sample size calculation for an unmatched case-control study
Cohort/RCT	Sample size for cross-sectional, cohort, clinical trial
Mean Difference	Sample size for the difference between two means
Power	<i>Power calculations for various study designs:</i>
Unmatched CC	Power calculations for an unmatched case-control study
Cohort	Power calculations for a cohort study
Clinical Trial	Power calculations for a randomized clinical trial
X-Sectional	Power calculations for a cross-sectional study
Mean Difference	Power calculations for the difference between two means
Random Numbers	Random number generator
Download OpenEpi	

- Two-by-two crude and stratified tables for count and rate data
- Matched case-control analysis
- Test for trend with count data
- Independent t-test and one-way analysis of variance (ANOVA)
- Diagnostic and screening test analyses with receiving operator characteristic (ROC) curves
- Sample size for proportions, cross-sectional surveys, unmatched case control, cohort, randomized controlled trials, and comparison of two means
- Power calculations for proportions (unmatched case control, cross-sectional, cohort, clinical trials) and for the comparison of two means
- Random number generator

For epidemiologists and other health researchers,

OpenEpi performs a number of calculations based on cross-tabulations not found in most epidemiologic and statistical programs. For example, for a single two-by-two table, in addition to the conventional results presented in other programs, OpenEpi provides estimates for:

- Etiologic or prevented fraction in the population and exposed with confidence intervals, based on risk, odds, or rate data
- The cross-product and maximum likelihood odds ratio estimate
- Mid-*p*, exact *p*-values, and confidence limits for the odds ratio
- Calculations of rate ratios and rate differences with confidence intervals and statistical tests

For stratified two-by-two tables with count data, OpenEpi provides:

- Mantel-Haenszel and precision-based estimates of the risk ratio and odds ratio
- Precision-based adjusted risk difference
- Tests for interaction for the risk ratio, odds ratio, and risk difference
- Four different confidence limit methods for the odds ratio

Similar to Epi Info, in a stratified analysis, both crude and adjusted estimates are provided so that the assessment of confounding can be made. With rate data, OpenEpi provides adjusted rate ratios and rate differences, and tests for interaction. Finally, with count data, OpenEpi also performs a test for trend, for both crude and stratified data.

An example of one of the modules is shown in Figure 2, which features five tabs. The “Start” tab provides an overview of the module. The “Enter” tab is where data entry is performed, or the user can click on the “Enter New Data” button. Results of the calculations are under the “Results” tab. A number of examples are provided under the “Examples” tab. And a “Help” tab is available to answer questions about the program. Other links on the screen include “Documentation,”

which provides the analytic details, including formulae, used in the module; and “Testing,” which compares the results of the OpenEpi module to other software programs or to textbook examples. Many of the modules have a “Load Demo Data” button that will automatically load example data into the module.

## USE OF OPENEPI IN TEACHING

To the developer’s knowledge, OpenEpi has been used as a training tool for teaching epidemiology to students at Columbia University, Emory University, Morehouse College, San Jose State University, University of Medicine and Dentistry of New Jersey, University of Michigan, and University of Wisconsin, for campus-based and distance-learning courses. Because OpenEpi is easy to use, requires no programming experience, and can be run on the Internet, students can use the program and focus on the interpretation of results.

## FUTURE DIRECTIONS FOR OPENEPI

While much has been accomplished, the investigators would like to enhance OpenEpi by (1) improving

Figure 2. Example of “Start” page for two-by-two table module

**OpenEpi** Start Enter Results Examples Help

Open Source Statistics for Public Health

**Enter New Data**

Add Stratum | Stratum 1 | Delete Stratum

**OpenEpi 2 x 2 Table**

		Exposure		Totals
		(+)	(-)	
Disease	(+)	36	36	102
	(-)	28	32	60
Totals		64	68	132

**Author(s)**

**Statistics**  
 Kevin M. Sullivan, Emory University  
 and Andrew G. Dean, EpiInformatics.com  
 based on code from John C. Pezzullo  
 Exact and maximum likelihood statistics adapted from  
 a Pascal program by David Martin. Thanks to Ray  
 Simons for advice and testing.

**Interface**  
 Andrew G. Dean and Roger Mir

**Two by Two Tables**

Two by two tables are used to evaluate the association between a possible risk factor ("Exposure") and an outcome ("Disease"). Counts summarizing the occurrence of the four possible combinations of events in the study population are entered into the appropriate cells. The table can be rotated or flipped so that either rows or columns represent Exposure, and the column headings (+) and (-) can be in either order to match common textbooks of epidemiology. A single table or multiple strata can be entered.

Statistics produced include the Fisher and mid-p exact tests, chi squares, odds ratio, maximum likelihood odds ratio estimate, risk/prevalence ratio (relative risk), risk difference, and etiologic fractions with confidence limits produced by several methods, with stratified analysis

**Load Demo Data**

Stratum	Measures of Association					
	Chi Square	p value	Odds Ratio	95% CI	Mid-P Exact	p value
1	5.047	0.0248	4.374	0.9756	5.046	0.02511
2	1.486	0.2231	1.228	0.2708	1.482	0.2239
All expected values from test/Probable and grand total are >= 5						
OK to use chi square						
Stratum	Chi Square	p value	Odds Ratio	95% CI	Mid-P Exact	p value
1	5.047	0.0248	4.374	0.9756	5.046	0.02511
2	1.486	0.2231	1.228	0.2708	1.482	0.2239
All expected values from test/Probable and grand total are >= 5						
OK to use chi square						
Stratum	Chi Square	p value	Odds Ratio	95% CI	Mid-P Exact	p value
1	5.047	0.0248	4.374	0.9756	5.046	0.02511
2	1.486	0.2231	1.228	0.2708	1.482	0.2239
All expected values from test/Probable and grand total are >= 5						
OK to use chi square						

Select, copy, and paste results to other programs or download OpenEpi to local disk and run OpenEpiSave HTA to save automatically.

the existing programs; (2) adding new programs to perform epidemiologic, statistical, or nutritional calculations; (3) adding the ability to cut and paste and read data; (4) translating the program into additional languages; and (5) increasing the level of knowledge of OpenEpi worldwide. A number of prototype OpenEpi modules, including logistic regression and survival analysis, can be found at <http://www.sph.emory.edu/~cdckms>. We invite others to develop and test new modules for OpenEpi.

---

Kevin Sullivan is an Associate Professor with the Department of Epidemiology, Rollins School of Public Health of Emory University, Atlanta, Georgia. Andrew Dean is an Adjunct Professor with the Department of Global Health, Rollins School of Public Health of Emory University. Minn Minn Soe is a Medical Epidemiologist with the Tennessee Department of Health, Nashville, Tennessee.

Address correspondence to: Kevin M. Sullivan, PhD, MPH, MHA, Rollins School of Public Health, Emory University, Grace Crum Rollins Building, 1518 Clifton Rd., Atlanta, GA 30322; tel. 404-727-8714; fax 404-727-8710; e-mail <[cdckms@sph.emory.edu](mailto:cdckms@sph.emory.edu)>.

## REFERENCES

1. Sullivan K. Google search identifies programs epidemiologists use most. *The Epidemiology Monitor* 2008;29:4.
2. Sullivan K, Abramson J. Update on free epidemiologic software. *The Epidemiology Monitor* 2008;29:3-4.
3. Sullivan KM. OpenEpi now available in English, French, Italian, and Spanish. *The Epidemiology Monitor* 2007;28:14.
4. Sullivan KM, Dean A, Soe MM. OpenEpi version 2. *The Epidemiology Monitor* 2007;28:1,9-10.
5. Sullivan KM. Overview of free analytic software for epidemiologists. *The Epidemiology Monitor* 2007;28:3-4.
6. Sullivan K, Dean A, Soe MM. OpenEpi: an update. *The Epidemiology Monitor* 2006;27:4.
7. Sullivan KM, Dean AG, Mir R. OpenEpi: a new collaborative effort in epidemiologic computing. *The Epidemiology Monitor* 2004;25:3,7,9.